Outcomes of Care

The Effect of Hospital Volume on the In-Hospital Complication Rate in Knee Replacement Patients

Edward C. Norton, Steven A. Garfinkel, Lisa J. McQuay, David A. Heck, James G. Wright, Robert Dittus, and Robert M. Lubitz

Objective. To examine the effect of hospital volume on in-hospital surgical outcomes for knee replacement using six years of Medicare claims data.
Data Sources/Study Setting. The data include inpatient claims for a 100 percent sample of Medicare patients who underwent primary knee replacement during 1985–1990. We supplemented these data with information from HCFA's denominator files, the Area Resource File, and the American Hospital Association survey files.
Study Design. We estimated the probability that a patient has an in-hospital complication in the initial hospitalization for the first primary knee replacement, using a Logit model, for three definitions of complication. The models controlled for hospital volume, other hospital characteristics, patient demographics, and patient health status. We tested for the endogeneity of hospital volume.
Data Collection/Extraction Methods. A panel of two orthopaedic surgeons and two internists reviewed diagnosis codes to determine whether a complication was likely, possible, or due to anemia. After removing the few observations with bad or missing data, the final population has 295,473 observations.
Principal Findings. The probability of a likely in-hospital complication declines rapidly from 53 through 107 operations per year, then levels off. Statistical tests imply that hospital volume is exogenous in this patient-level data. Complication rates increased steadily through the study period. Although obesity appeared to lower the probability of a complication, a counterintuitive result, further investigation revealed this to be an artifact of the claims data limit of listing no more than five diagnoses. Controlling for this restriction reversed the effect of obesity.
Conclusions. Rather than uncontrolled expansion of knee surgery to small hospitals, decentralization to regional centers where at least about 50, and preferably about 100, operations per year are assured appears to be the optimal policy to reduce in-hospital complications.

Key Words. Hospital volume, knee replacement, outcomes research

Knee replacement has become one of the most common inpatient surgical procedures among the elderly, with nearly 80,000 performed on Medicare patients in 1990 (Coyte et al. 1995). The number of knee replacements in 1990 was almost double those performed in 1985. The large number of knee replacements leads to the question of what factors influence a good outcome. Recent studies by the Total Knee Replacement Patient Outcomes Research Team (TKR PORT) have focused on outcomes defined as improvements in function and pain relief (Callahan et al. 1994; Hawker, Bombardier, Freund, et al. 1993); prosthesis failure (Callahan et al. 1994; Heck, Melfi, Mamlin, et al. 1997); selected complications (e.g., thromboembolism) (Callahan et al. 1994); and mortality (Dittus, Katz, Heck, et al. 1995). These studies have found knee replacement surgery to be effective. Because of the importance of hospital volume for public policy, this article focuses on the effect of hospital knee replacement volume on the immediate postoperative complication rate during the initial hospitalization, which we call the in-hospital complication rate. If hospitals with higher volume have lower in-hospital complication rates, then there is a policy trade-off between making knee replacement available at more hospitals to improve access (Coyte, Wright, Hawker, et al. 1994) or making knee replacements available at fewer hospitals to improve the quality of care.

Numerous studies have shown that hospitals with higher surgical volume have lower mortality rates (Luft, Bunker, and Enthoven 1979; Luft 1980; Luft, Hunt, and Maerki 1987; Hannan, Kilburn, Bernard, et al. 1991; Hosenpud, Breen, Edwards, et al. 1994; Hughes, Garnick, Luft, et al. 1988; Fowles, Bunker, Oda, et al. 1988; Hannan, Siu, Kumar, et al. 1995; Phillips, Luft, and Ritchie 1995; Roos, Black, Roos, et al. 1995). Furthermore, these

---

This research was funded by the U.S. Agency for Health Care Policy and Research under grant No. 06432 to establish a Patient Outcomes Research Team (PORT) to study Total Knee Replacement (TKR) surgery. James Wright acknowledges support of a MRC scholarship.

Address correspondence and requests for reprints to Edward C. Norton, Ph.D., Assistant Professor, Department of Health Policy and Administration, School of Public Health, CB #7400, McGavran-Greenberg Hall, University of North Carolina, Chapel Hill, NC 27599-7400. Steven A. Garfinkel, Ph.D. is Senior Health Services Analyst, Research Triangle Institute, Lisa J. McQuay, is Senior Systems Analyst, Glaxo Wellcome; David A. Heck, M.D. is Professor of Orthopaedic Surgery, Indiana University; James G. Wright, M.D., M.P.H. is Associate Professor of Surgery and Public Health Sciences, University of Toronto; Robert Dittus, M.D., M.P.H. is Professor of Investigative Medicine, Vanderbilt University; and Robert M. Lubitz, M.D., M.P.H. is Clinical Associate Professor, St. Vincent's Hospital, Indianapolis, IN. This article, submitted to *Health Services Research* on August 13, 1996, was revised and accepted for publication on February 23, 1998.

studies suggest that there are diminishing returns to volume, meaning that a given increase in volume has a greater beneficial effect on outcome for a small-volume hospital than for a large-volume hospital. Most of these studies have focused almost exclusively on cardiac surgery, which is potentially life-saving, and have not studied surgery that enhances quality of life, but is not life-saving, such as knee replacement. The study by Fowles and colleagues of total hip replacement is an exception.

Studies of hospital surgical volume have typically looked only at mortality as the outcome of interest, while ignoring other complications. However, information is limited on the effect of orthopaedic interventions on other outcomes. In our Medicare study population, 30-day postsurgical mortality for knee replacement was only 0.63 percent (Dittus, Katz, Heck, et al. 1995). Thus, mortality alone is insufficient for assessing the relationship between volume and outcome because death is too infrequent. Instead, this study focused on a broad range of complications, including mortality. This study is one of the few to evaluate the effect of surgical volume on outcomes other than mortality (for others see Fowles, Bunker, Oda, et al. 1988; Phillips, Luft, and Ritchie 1995; Roos, Black, Roos, et al. 1995). However, the study by Fowles and colleagues did provide insights into another orthopaedic intervention, total hip replacement.

This study has five other methodological improvements over prior research. First, we conduct more extensive tests for endogeneity of hospital volume and outcomes than described elsewhere in the literature. Second, we show that limiting the diagnostic fields to five may lead to biased results on the effects of obesity. Third, we demonstrate the importance of treating anemia separately from other in-hospital outcomes. Fourth, we control for the correlation of patients within the same hospital. Fifth, we show that claims data are valuable for morbidity-related outcomes studies because their low cost relative to medical record abstracts means that we could include variation in hospital characteristics in our models, make inferences to a national population, and perform extensive tests for endogeneity.

Our study population included all Medicare beneficiaries (except for a few excluded groups) who had knee replacement surgery from 1985 through 1990. We obtained hospital characteristics from the American Hospital Association and patient characteristics from the Medicare enrollment and claims files. A panel of four physicians designated the in-hospital complications based on diagnostic codes from the Medicare claims files using a Delphi process. We used alternative definitions to establish the sensitivity and robustness of our results.

METHODS

We tested the hypothesis that hospitals where more knee replacements are performed have better outcomes. Our model measured the effect of volume on the probability of having an in-hospital complication from knee replacement surgery controlling for other risk factors. We defined volume as the total number of knee replacements for Medicare beneficiaries performed at a hospital in one year. The relation between volume and the probability of a complication may be nonlinear. For example, if there are diminishing returns to volume, then each successive operation has a smaller beneficial effect on the probability of a complication and a linear model would yield biased results. Our model used a piecewise linear specification to allow for a flexible relationship between volume and the probability of a complication.

The probability of a complication also depends on other hospital and patient characteristics, such as teaching status and occupancy rate, which may affect the quality of the staff and their willingness to accept high-risk patients. Patients' age and comorbidities may also affect the probability of a complication. Medical knowledge may improve over time, so we also expect there to be a time trend. The general relationship posed by our model is described as follows:

$$\Pr(\text{Complication}_{ijt}) = f(\text{Hospital}_{jt}, \text{Patient}_i, \text{Time trend}_t) \qquad (1)$$

where f(•) is some function, *i* is the subscript for patients, *j* is the subscript for hospitals, and *t* is the subscript for year.

We estimated the model for three different measures of complications for two reasons. First, the use of multiple outcome measures reflects the difficulty in designating complications from ICD-9-CM diagnostic codes. Some codes can be designated a complication with more certainty than others, and the alternative measures reflect the difference. Second, multiple outcomes enable us to test the sensitivity of our results to the definition of complications.

We also tested the hypothesis that hospital volume is endogenous. Other researchers, *using hospital-level data*, have found that hospitals with better outcomes attract more patients (Luft 1980; Luft, Hunt, and Maerki 1987; Hughes, Garnick, Luft, et al. 1988). We use *patient-level data* for our analysis, which, we hypothesized, should not be endogenous. Although hospitals with low complication rates should attract more patients, the probability of a single individual having a complication should not have a measurable effect on hospital volume. If hospital volume is endogenous, then ignoring the endogeneity will bias the coefficient on volume in the complication equation.

However, if volume is not endogenous, then using instrumental variables will increase the variance of the estimator. We used standard statistical tests to decide whether the decrease in bias is worth the increase in variance.

We rigorously tested for endogeneity in a simultaneous equations model in which the first-stage regression predicted hospital volume using ordinary least squares, and the main equation used a Logit model to predict whether a patient had a likely complication (Bollen, Guilkey, and Mroz 1995). Following the literature, the instruments were dummy variables measuring the number of beds in the hospital (Luft 1980; Luft, Hunt, and Maerki 1987; Hughes, Garnick, Luft, et al. 1988). We used five dummy variables for five ranges of number of beds. The results showed no evidence of endogeneity, based on four tests. First, we tested whether volume is exogenous by including in the main equation both the actual volume and the estimated error from the first-stage volume regression. The coefficient on the error term when included in the main equation is not statistically significant (t-statistic = 1.80). Therefore, we do not reject the hypothesis that volume is exogenous. Second, we tested whether the instruments were significantly correlated with the potentially endogenous variable. A significant correlation is required for the model to be identified. The t-statistics on the four instruments in the first-stage regression were all significant, ranging in absolute value from 5.3 to 14.9. The F-test of their joint significance was significant at the one percent level. Therefore the model is identified. Third, the explanatory power of the first-stage regression should be large so that the measure of volume is not replaced with a noisy measure. We tested for explanatory power by measuring the R-squared in the first-stage regression, which was 0.42. This value of R-squared is considered quite high (Bollen, Guilkey, and Mroz 1995). Fourth, we tried a wide variety of specifications of both volume and instrumental variables to test for the robustness of the results. The results of the endogeneity tests were qualitatively similar. We conclude that correcting for any potential bias by using simultaneous methods is not worth the increase in variance.

Our finding of no endogeneity is in contrast to that of the researchers who used hospital-level data. This conclusion is not surprising. The endogeneity between hospital volume and hospital complication rates is at the hospital level, not the patient level.

DATA

The Health Care Financing Administration's Medicare Provider Analysis and Review (MEDPAR) files provided the principal data for this study. These files contain data on 100 percent of Medicare-reimbursed hospitalizations, except those delivered to beneficiaries enrolled in Medicare HMOs. We supplemented these data with information from the Health Care Financing Administration's denominator finder files, the Area Resource File, and the American Hospital Association survey files.

We drew the original sample by searching Medicare Part A inpatient claims from calendar years 1985 through 1990 to identify all hospitalizations in which a patient underwent primary knee replacement surgery based on the presence of ICD-9-CM procedure codes 81.41, 81.54, and 81.55. Medicare claims are reliable for identifying procedures such as knee replacement surgery (Paul, Melfi, and Kalasinski 1994; Mitchell, Bubolz, Paul, et al. 1994; Iezzoni, Daley, Heeren, et al. 1994). To make the data as complete as possible and to exclude hospitalizations during which it was highly questionable that a knee replacement actually occurred, we removed several types of patients and claims from the data set. We excluded patients enrolled in a health maintenance organization, because Medicare claims are not usually filed for HMO enrollees; patients whose health status was presumed to be very different from that of other Medicare patients (because they are under age 65 and currently eligible for Medicare due to a disability, or because they have end-stage renal disease, even if older than age 65); patients who resided outside the United States, Puerto Rico, and the Virgin Islands; patients who had a diagnosis code indicating that the procedure was not done because of contraindication or patient preference (code V64); patients for whom it appeared certain that the procedure was miscoded due to an unusually short length of stay and low charges; patients who received the surgery in an inappropriate place, such as a psychiatric, rehabilitation, or drug treatment facility; patients admitted to a hospital that could not be matched to the American Hospital Association data and those with missing or invalid patient-level variables; and any claim in which the procedure appeared to be a reoperation after a previous knee replacement (Heck, Melfi, Mamlin, et al. 1997). The final population of primary knee replacements has 295,473 observations. Each observation represents a primary knee replacement surgery. For patients with more than one primary knee replacement, only the record with the earlier date of service was retained.

The dependent variable of interest is an indicator of whether or not the patient had an in-hospital complication. We used the claims data to identify diagnoses that were likely to indicate a complication, including mortality. To define complications as accurately as possible, we listed diagnosis codes in descending order of frequency of their appearance in the claims, regardless of position on the claim (primary through fifth). Only diagnoses that were coded in less than 0.5 percent of all code entries in the data set were excluded from further consideration. The remaining diagnosis codes were reviewed by a panel of two orthopaedic surgeons and two internists using a Delphi process to decide which were complications. All four physicians agreed on the likelihood complication strategy employed as well as on the assignment of codes to the designated category. The four most common diagnosis codes for likely complication were disorders of fluid, electrolyte, and acid-base balance; cardiac dysrhythmias; complications affecting specified body systems; and rupture of tendon. The final set of codes judged to be complications are available from the authors upon request.

We divided diagnostic codes into those considered to be a likely complication and those considered to be a possible complication. By our definition, the set of possible complications includes all likely complications. About 17 percent of our sample had a likely complication, including the few who died in the hospital and the few who had a revision of their primary knee replacement. About 33 percent of our sample had a possible complication according to this definition.

Anemia—by far the most frequently used code except for the arthritis codes—was analyzed separately as a complication. On the one hand, mild anemia is an expected outcome of surgery with few sequelae and, therefore, may not be a complication. On the other hand, for some patients, the anemia code may have indicated severe, unexpected anemia. Thus, we treated anemia as a separate outcome variable. Overall, about 11 percent of the sample had anemia as a diagnosis.

From the MEDPAR data we computed the explanatory variable of primary interest—hospital volume—as the number of knee replacement procedures for Medicare beneficiaries performed at each hospital in each calendar year, excluding those for HMO enrollees and beneficiaries with end-stage renal disease. Therefore, the measure of volume includes reoperations and knee replacements for persons with missing data for other variables. The average volume per year in our sample is about 63 per hospital. The distribution of volume per hospital per year is highly skewed: in 1985, the first quartile, median, third quartile, and maximum were 4, 10, 22, and 313; by 1990 these numbers had risen to 7, 18, 39, and 444. To allow for a nonlinear effect of hospital volume, we recoded volume as a set of four splines, with cutoff points of 20, 40, and 80 knee replacement operations per year. Roughly a quarter of all patient observations fall into each category.

We also included a dummy variable for whether the hospital was new to knee replacement surgery, defined as having done no operations in the prior year.

Although our measure of volume excludes knee replacements for the non-Medicare population, about three-quarters of knee replacement patients are age 65 or older (Graves 1995). We assume that hospitals that perform more knee replacements on Medicare patients also perform more knee replacements on other patients. The standard errors and significance levels are unbiased under the assumption that the fraction of Medicare patients is the same across hospitals. We checked the fraction of hospital discharges attributable to Medicare beneficiaries for every hospital in our sample and were surprised by the small variation. The majority of hospitals have between 30 and 41 percent of their inpatient discharges paid by Medicare, and the distribution has a very sharp peak around 35 percent. Hospitals at the tenth and ninetieth percentiles have 24 percent and 48 percent of inpatient discharges covered by Medicare. To the extent that we mismeasure the true volume, the parameter estimates for volume are biased toward 0 by the usual argument about measurement error. This makes our results even stronger.

We divide the remaining explanatory variables into hospital characteristics and patient characteristics. The hospital characteristics that may affect the complication rate are divided into three groups: type of hospital, type of clientele, and services offered. Controlling for hospital type is important because hospitals specializing in knee replacement or the training of orthopaedic surgeons will probably have more severe and complex cases. We include measures of whether the hospital is for-profit or government, with the baseline category being nonprofit. We define variables for whether the hospital is an orthopaedic or other specialty hospital using the AHA designation and a teaching hospital by membership in the Council of Teaching Hospitals. We also control for whether the hospital is a member of a multihospital system. Hospitals in multihospital systems may be able to share information and use the experience of other hospitals in the system. Just over one-third of the patients visit hospitals in a rural area.

For clientele, we include the percentage of hospital inpatient days for Medicare patients and the occupancy rate. About 47 percent of inpatient days are covered by Medicare, and the average occupancy rate is 66 percent. To control for the type of services offered, we include dummy variables for the presence of physical therapy and outpatient rehabilitation services. We hypothesize that hospitals with physical therapy and outpatient rehabilitation services will attract more complex cases and that the probability of complications will decrease as the overall frequency of operations increases. We also include dummy variables indicating which census region the hospital is in to capture broad differences in practice style.

Hospital data from the annual American Hospital Association surveys are matched to patients by hospital and by year of discharge to reflect any changes that may have occurred over time. Some hospital characteristics vary by year, especially the primary explanatory variable, volume. Therefore, patients who go to the same hospital will have identical hospital variables only if they are discharged in the same year.

Our model included a dummy variable for each year after 1985 to control for any time trend in changes in overall complication rates. The number of patients who had a knee replacement increased each year from 1985 through 1990, so we expect that the characteristics of patients receiving knee replacement and providers performing knee replacements may have changed during this period. A significant positive effect of the year indicators would indicate that the probability of complication has increased relative to 1985. This result would be consistent with either riskier patients or with less proficient providers participating in knee replacement as it becomes more common. Changes in coding practice could also affect the baseline probability of a complication. A significant negative effect would be consistent with providers becoming more skilled, improvements in technology, or healthier patients.

Patient characteristics include age, gender, race, and comorbidities for case-mix adjustment. Age is measured as true age less 65 so that the intercept refers to a 65-year-old person. The average age is 74 (see the last column of Table 1). About one-third of the patients are men. Over 90 percent are white, 5 percent are African American, and one percent are neither African American nor white.

The remaining patient-level dummy variables control for case-mix severity. There is a variable for whether the patient was admitted through the emergency room and another variable for whether the patient was originally eligible for Medicare because of a disability (almost 9 percent of the sample). Both measures indicate poor health status compared with the average Medicare patient. We control for whether the patient is clinically obese since obesity may hinder both the operation and the recovery. We also interact obesity with gender to see whether obesity has a differential effect on men and women. A dislocated knee, quite rare in our sample, may also be more likely to result in a complication from surgery. Finally, 45 diagnoses that were judged unlikely to be complications are used as case-mix adjusters. We predicted that most of these case-mix control variables would increase the probability of a complication.

Table 1: Logit Results for Predicting Probability of Complication

Variable	*Likely Complication*	*Possible Complication*	*Anemia*	*Mean*
Constant	−1.97**	−0.831**	−2.97**	1.0
	(0.11)	(0.096)	(0.25)	
Volume Splines				
0–20	−0.0008	−0.0030	0.0048	18.1
	(0.0022)	(0.0018)	(0.0038)	[range 1–20]
21–40	0.0014	−0.0022	−0.0099**	12.9
	(0.0017)	(0.0014)	(0.0033)	[range 0–20]
41–80	−0.0047**	−0.00250**	−0.0024	14.9
	(0.0012)	(0.00094)	(0.0024)	[range 0–40]
>80	0.00063	0.00042	−0.00091	16.9
	(0.00057)	(0.00045)	(0.00097)	[range 0–364]
Volume = 0 in prior year	0.076	0.055	0.18	0.0079
	(0.078)	(0.057)	(0.11)	
Hospital Characteristics				
For-profit	0.186**	0.163**	0.320**	0.093
	(0.041)	(0.034)	(0.070)	
Government	0.061	0.061*	0.111	0.097
	(0.035)	(0.028)	(0.067)	
Orthopaedic	0.41*	0.31*	0.55*	0.019
	(0.17)	(0.14)	(0.25)	
Other specialty	0.89*	0.59*	1.10	0.003
	(0.41)	(0.30)	(0.58)	
Teaching	0.102**	0.088**	−0.375**	0.163
	(0.037)	(0.028)	(0.077)	
Multihospital system	0.002	0.021	−0.050*	0.459
	(0.024)	(0.019)	(0.047)	
Patient Characteristics				
Age minus 65	0.02410**	0.02657**	0.0210**	8.72
	(0.00089)	(0.00073)	(0.0012)	
Male	0.261**	0.0183	−0.283**	0.327
	(0.011)	(0.0094)	(0.016)	
African American	0.113**	0.202**	0.181**	0.046
	(0.026)	(0.021)	(0.037)	
Other race	−0.015	−0.037	0.008	0.010
	(0.050)	(0.041)	(0.063)	
Emergency admission	0.170**	0.516**	0.014	0.0095
	(0.054)	(0.051)	(0.087)	
Disability	0.138**	0.188**	0.052*	0.088
	(0.018)	(0.014)	(0.024)	
Obese	−0.286**	−0.133**	−0.194**	0.037
	(0.038)	(0.026)	(0.049)	
Obese and male	−0.024	−0.014	−0.15	0.0058
	(0.080)	(0.062)	(0.11)	
Dislocated knee	−0.41	0.76**	−1.18**	0.0006
	(0.23)	(0.16)	(0.39)	

Continued

Table 1: Continued

Variable	*Likely Complication*	*Possible Complication*	*Anemia*	*Mean*
Include patient comorbidities	Yes	Yes	Yes	
Include region dummy variables	Yes	Yes	Yes	
N	295,473	295,473	295,473	

* Statistically significant at the 5 percent level; ** statistically significant at the 1 percent level.

STATISTICAL METHODS

We estimated the probability that a patient has a complication following the first primary knee replacement using a Logit model. We ran Logit models for each of the three types of complications: likely, possible, and anemia. A positive coefficient indicates that an increase in the corresponding covariate will increase the probability that the patient has a complication.

Patients treated in the same hospital tend to be similar to each other in ways that are unobservable. For example, hospitals differ in the skill and training of the physicians. Patients who go to the same hospital for a knee replacement will likely be treated by the same group of physicians. Such correlation due to clustering at the hospital level may bias the standard errors, typically in a way that exaggerates the significance of the results (Norton et al. 1996a). We corrected the standard errors using generalized estimating equations, also known as the Huber correction, with clustering at the hospital-year level. This correction did not affect the coefficients, but the standard errors increased roughly by a factor of two for hospital characteristics, and by a few percent for patient characteristics. All reported results have the robust standard errors. Other studies of the effect of volume on outcome that analyze patient-level data have not controlled for correlation of patients within hospitals.

RESULTS

Our results showed a strong beneficial effect of higher volume on the probability of a likely complication (see the first column of Table 1). The coefficient on the third volume spline was negative and highly significant, implying that the probability of a likely complication decreases when between 40 and 80 operations are performed per year, and then levels off. Therefore, the marginal effect of volume was nonlinear—greatest between 40 and 80 operations per year and close to zero below 40 and above 80. The results imply that increasing the number of operations from 40 to 80 would decrease the probability of a likely complication by about 3 percentage points, compared to an overall mean of 17 percent.

A similar pattern was found for the effect of volume on the probability of a possible complication (see the second column of Table 1). The coefficient on volume between 40 and 80 was also negative and highly statistically significant ($p < .01$). For anemia, the pattern is slightly different. The coefficient on volume between 20 and 40 is negative and highly statistically significant ($p < .01$). Hospitals that were performing knee replacement operations for the first year had no different probability of a complication.

Five hospital characteristics were generally significant. Patients in for-profit, orthopaedic, and other specialty hospitals were much more likely to have complications. Patients in teaching hospitals and hospitals with high occupancy rates were more likely to have a likely or possible complication, but less likely to have anemia.

The remaining hospital characteristics did not have a consistent statistically significant effect. Government hospitals, rural hospitals, and hospitals that are part of multihospital systems were neither more nor less likely to have patients with complications. The presence of physical therapy services in the hospital decreased the probability of all three types of complications, but these results were not statistically significant. In contrast, the presence of outpatient rehabilitation services in the hospital increased the probability of all three types of complications. It was not clear why hospital characteristics should have a different effect depending on the type of complication. However, mild anemia is considered by some physicians to be a natural and expected consequence of knee surgery; thus, the difference in result for anemia compared with other complications may be the result of coding practices.

The strong time trend shown in the summary statistics persisted in the multivariate analyses. Each year the probability of a complication rose significantly. The annual increase in the probability of each type of complication was almost constant. Each year the probability of a likely complication increased by about 2 percent, and of a possible complication by about 3 percent. The probability of anemia rose at an increasing rate.

We also ran one further set of regressions to test whether the general results held up for each year. We reran the models presented in Table 1 with interaction terms between the volume splines and the year dummy variables. The results indicate whether the relationship between volume and the complication rate is different in each year when compared with the base year of 1985. The coefficients on these interaction terms were not statistically significant in 19 out of 20 cases at the 5 percent level. Therefore, no difference between 1985 and the other years appeared in the relationship.

The coefficients on the patient characteristics in general had the predicted sign and were statistically significant. Older patients were more likely to have a complication, as were African Americans. Men were more likely than women to have a complication, but women were more likely than men to have anemia. Other race had no significant effect. Being admitted for an emergency and being originally eligible for Medicare through disability increased the probabilities of all complications. Having a dislocated knee had no significant effect on a likely complication, but had a positive effect on a possible complication and a negative effect on anemia.

The one surprising result was that being obese decreased the probability of all three kinds of complications. We had predicted that an obese person would be more likely to have complications both because of the increased difficulty for the surgeon in performing the surgery and of rehabilitation. Another TKR PORT study found that obese patients are more likely to have urologic complications and longer time in surgery by 46 minutes (Lubitz, Dittus, Paul, et al. 1995). We found that the contrary obesity result was an artifact of the limited number of diagnostic codes than could be entered on the claims form used. Each patient had up to five diagnoses listed on his or her claims form. The patients' primary diagnosis was related to knee replacement in this sample, which filled one diagnosis field. Most patients had several other diagnoses, so that even if the patient was obese it might not have been recorded as a diagnosis due to lack of space. If the patient was both obese and had a complication, then the complication diagnosis might have crowded out the obesity diagnosis. In other words, it was more likely that obese patients would have been listed as obese if they did not have a complication. To test this hypothesis we reran the full models for all three types of complications using a restricted sample (Norton et al. 1996b). We kept all patients listed as obese and all patients who had fewer than five diagnoses in the sample. We dropped all patients who were not listed as obese but for whom there was no room to list obesity, leaving a sample size of 230,195. The results confirmed our hypothesis. The sign on obesity was positive instead of negative, and was statistically significant at the one percent level in all three regressions using the limited sample (see Table 2 for the results of the likely complication). The signs and magnitudes of the other important variables remained nearly unchanged.

Table 2: Logit Results for Obesity

Variable	*Likely Complication*
Constant	−2.29**
	(0.13)
Volume Splines	
0–20	0.0004
	(0.0027)
21–40	0.0009
	(0.0019)
41–80	−0.0065**
	(0.0012)
>80	0.00117
	(0.00060)
Volume = 0 in prior year	0.147
	(0.086)
Patient Characteristics	
Obese	0.461**
	(0.039)
Obese and male	−0.064
	(0.081)
Include other hospital characteristics	Yes
Include year dummy variables	Yes
Include other patient characteristics	Yes
Include patient comorbidities	Yes
Include region dummy variables	Yes
N	230,195

* Statistically significant at the 5 percent level; ** statistically significant at the 1 percent level.

DISCUSSION

Our study has produced a number of interesting methodological, policy-relevant, and clinical findings. In the methodological area, we have four notable conclusions. First, we have conducted extensive tests for endogeneity in the relationship between knee replacement and outcomes (as defined by subsequent morbidity) and have found that the relationship is not endogenous. This finding is a significant contribution to the literature on volume-outcomes because (1) we have performed more extensive testing than is described elsewhere in the literature and (2) we have demonstrated that individual beneficiary- or patient-level data are superior to aggregated hospital-level data for studying the volume-outcome relationship. At the aggregated hospital level, volume is as likely to be influenced by outcomes as is the reverse. However, based on extensive statistical testing, we conclude that this endogeneity does not hold for individual patients. Thus, patient-level data should be used whenever possible to study the relationship between volume and outcomes.

Second, we have demonstrated the effect of a limited number of diagnostic coding fields on the ability to study or control for the effects of obesity in research on the outcomes of elective surgical procedures (Romano and Luft 1992). In knee replacement surgery, obesity as a comorbidity was crowded out often enough to reverse the sign on the coefficient and yield a significant counterintuitive finding that was entirely an artifact of coding. When we tested for this artifact by dropping the cases with all five diagnostic codes present but none coded for obesity, the sign reversed, yielding the expected significant relationship without changing any of the other signs or coefficients in a meaningful way. In econometric terms, the coefficient for obesity is biased to being negative because being obese is endogenous to having a complication. This was a remarkably strong demonstration of the effect of limiting diagnostic fields to five.

Third, we demonstrated the importance of treating anemia as a separate outcome in the study of volume and morbidity outcomes for elective surgery. We found several instances in which the results were reversed for likely complications with anemia. For example, patients in teaching hospitals were more apt to have the likely complications but less apt to have anemia than were other patients. Combining anemia with other likely complications would have their distinct effects.

Fourth, we demonstrated the value of claims data for morbidity-related outcomes studies. Because claims data are less expensive to obtain than medical record abstracts for large populations, we were able to include the variation in hospital characteristics in our models and to make inferences to a national population—two advantages that are unavailable for studies based on medical records. The large number of observations gave us sufficient power to perform extensive tests for endogeneity and to control optimally for obesity by dropping cases that might have been subject to the coding artifact. Further, the large sample size allowed for a broad range of complications, including those that are rare.

In the area of policy, we also have four noteworthy findings. First and foremost, we confirmed the existence of the classic volume-outcomes relationship for knee replacement surgery. Patients at hospitals that do more operations have better outcomes than patients at hospitals that do fewer.

We found this relationship using in-hospital complications and in-hospital mortality, rather than mortality alone, and we found it consistently for three distinct morbidity-based definitions of complications over each of six years, from 1985 through 1990.

Second, we identified the specific threshold levels of operations at which the benefits of performing more operations begin and end. There is no benefit from performing more knee replacements until at least 40 operations are performed each year and there is no further benefit of performing more once 80 are being performed. Since these results are based on data from Medicare patients only, and about three-quarters of knee replacements patients are age 65 or older, the thresholds for total volume including the noneldelry are about one-third larger: 53 and 107.

Third, we observed the rates of complications increasing consistently over time, along with the numbers of knee replacements. The presence of such a strong time trend even after controlling for patient and hospital characteristics could reflect a change in coding practices due to the prospective payment system (PPS). Medicare implemented the PPS just before the beginning of our study. Two DRGs are related to knee surgery: one with complications and one without. The DRG with complications had a reimbursement rate in 1990 about 74 percent higher than the one without complications. Thus, hospitals have an incentive to code more diagnoses under PPS because Medicare reimbursement is linked to coded diagnoses. We find a slight increase over time in the number of diagnoses that are comorbidities. In 1985, 47 percent of our sample had only one comorbidity, 30 percent had two, 16 percent had three, 6 percent had four, and one percent had five. By 1990, 38 percent of our sample had only one comorbidity, 32 percent had two, 20 percent had three, 8 percent had four, and 2 percent had five. This finding could also reflect improved documentation, or the extension of knee replacement to riskier patients, or the introduction of less skilled surgeons into the pool of those performing knee replacement as demand increases. Although we cannot test these possible explanations with our data, Carter, Newhouse, and Relles (1990) found that only one-third of the change in the Medicare Case Mix Index from 1986 through 1987 was attributable to changes in the completeness of coding. Therefore, at least part of the increase in complications probably results from changes in the patient or surgeon pools.

Fourth—and troubling—is our finding that, after controlling for comorbidities, for-profit hospitals have higher complication rates than other hospitals. It makes sense that orthopaedic and other specialty hospitals experience higher complications rates because they accept patients who are riskier in some way that is not controlled for by our case-mix adjustments. However, it is unlikely that the same can be said for for-profit hospitals, leaving us with the conclusion that patients who have their knee replacements performed in for-profit hospitals have poor outcomes compared with those who receive them in non-profit hospitals. Given the shift of hospitals toward for-profit status or a for-profit management style, this finding suggests that we will see a continuing increase in complication rates in the future.

Our principal finding for clinical practice is that, as expected, obese patients are at greater risk of in-hospital complications than those who are not obese.

The results should be understood in light of the limitations of the study. First, we have only inpatient claims data for our population. Thus, our outcome complications measures are limited to those that occurred during the inpatient hospital stay, and we are unable to identify complications following discharge or to relate hospital volume to long-term complications. Nevertheless, in-hospital complications, such as death, infection, and stroke, are clinically important outcomes. Second, the findings depend on a lack of correlation between hospital volume and unobserved patient case mix (as they do in any case mix–adjusted model). The results will be biased, however, only if the omitted case-mix measures are correlated with the hospital's volume, after taking into account the 45 patient comorbidities and other patient characteristics. Third, it was impossible to measure the surgeon volume instead of the hospital volume because the medical provider identification number in our data does not identify a unique surgeon. Some surgeons have multiple Medicare provider numbers and some Medicare provider numbers reflect the experience of multiple surgeons. We therefore could not obtain linkage between individual surgeons and individual cases. Prior research by Fowles and colleagues shows that the relationship between surgeon volume and outcome is qualitatively the same as hospital volume and outcome for total hip replacement (Fowles, Bunker, Oda, et al. 1988). Surgeon volume may be a better predictor of mortality, but both measures predict other complications equally well. Finally, there was no validation of the complications and comorbidity codes through medical record review. However, the knee replacement codes were validated through either patient survey or medical record abstract for a random sample of 1,750 persons in the claims data set and were found to be accurate in 97.8 percent of the cases (Katz, Freund, Heck, et al. 1996).

CONCLUSION

Several studies from the TKR PORT have demonstrated the positive value of knee replacement surgery for patients in terms of reduced pain, improved function, and patient satisfaction (Callahan et al. 1994, 1995; Hawker, Wright, Coyte, et al. 1997; Hawker, Wright, Coyte, et al. forthcoming). These studies argue for increasing access to knee replacement surgery.

In identifying about 50–100 operations as the minimum range of operations per year that hospitals should perform to minimize in-hospital complication rates, we have provided a clear guide for policy designed to increase access. Rather than uncontrolled expansion of knee surgery to increasingly smaller hospitals, decentralization to regional centers where at least 50, and preferably 100 operations per year are assured appears to be the optimal policy to reduce in-hospital complications.

REFERENCES

Bollen, K. A., D. K. Guilkey, and T. A. Mroz. 1995. "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia." *Demography* 32 (1): 111–31.

Callahan, C. M., B. G. Drake, D. A. Heck, and R. S. Dittus. 1994. "Patient Outcomes Following Tricompartmental Total Knee Replacement: A Meta-Analysis." *Journal of the American Medical Association* 271 (17): 1349–57.

———. 1995. "Patient Outcomes Following Unicompartmental or Bicompartmental Knee Arthroplasty: A Meta-Analysis." *Journal of Arthroplasty* 10 (2): 141–49.

Carter, G. M., J. P. Newhouse, and D. A. Relles. 1990. "How Much Change in the Case Mix Index Is DRG Creep?" *Journal of Health Economics* 9 (4): 411–28.

Coyte, P. C., G. Hawker, C. Ricketts, and J. G. Wright. 1995. "Variations in Rheumatologists' and Family Physicians' Perceptions of the Outcomes and Indications for Knee Replacement Surgery." Working Paper. University of Toronto.

Coyte, P. C., J. G. Wright, G. A. Hawker, C. Bombardier, R. S. Dittus, J. E. Paul, D. A. Freund, and E. Ho. 1994. "Waiting Times for Knee-Replacement Surgery in the United States and Ontario." *The New England Journal of Medicine* 331 (16): 1068–71.

Dittus, R. S., B. P. Katz, D. A. Heck, E. Boatright, G. Westmoreland, R. M. Lubitz, D. A. Freund, P. C. Coyte, J. G. Wright, J. E. Paul, D. Z. Joyce, and W. W. Young. 1995. "Determinants of 30-day Mortality Associated with Knee Arthroplasty." Working Paper. Indiana University.

Fowles, J., J. P. Bunker, M. Oda, D. J. Schurman, and M. Loftus. 1988. "Relation of Surgical Volume to Outcomes and Charges: Pilot Study of Total Hip Replacement Using Northern California Medicare Data." In *Lessons from the First Twenty*

Years of Medicare, edited by M. V. Pauly and W. L. Kissick. Philadelphia, PA: University of Pennsylvania Press.

Graves, E. J. 1995. *Detailed Diagnoses and Procedures: National Hospital Discharge Survey, 1993.* National Center for Health Statistics. *Vital and Health Statistics* 13 (122).

Hannan, E. L., A. L. Siu, D. Kumar, H. Kilburn, Jr., and M. R. Chassin. 1995. "The Decline in Coronary Artery Bypass Graft Surgery Mortality in New York State: The Role of Surgeon Volume." *Journal of the American Medical Association* 273 (3): 209–13.

Hannan, E. L., H. Kilburn, Jr., H. Bernard, J. F. O'Donnell, G. Lukacik, and E. P. Shields. 1991. "Coronary Artery Bypass Surgery: The Relationship Between In-Hospital Mortality Rate and Surgical Volume After Controlling for Clinical Risk Factors." *Medical Care* 29 (11): 1094–107.

Hawker, G. A., C. Bombardier, D. Freund, R. S. Dittus, D. Heck, P. Coyte, J. G. Wright, B. Katz, and J. E. Paul. 1993. "U.S. Survey of Knee Replacement Recipients to Assess Short and Long-Term Outcomes." *Arthritis and Rheumatism* 36 (Supplement): S85.

Hawker, G. A., J. G. Wright, P. C. Coyte, J. E. Paul, R. S. Dittus, R. Croxford, and B. P. Katz. 1997. "Health-related Quality of Life After Knee Replacement Surgery." Working Paper. University of Toronto.

Hawker, G. A., J. G. Wright, P. C. Coyte, R. Croxford, J. E. Paul, B. P. Katz, C. Bombardier, C. A. Melfi, and R. S. Dittus. Forthcoming. "Patient-Reported Knee Pain Following Knee Replacement Surgery." *Journal of Bone and Joint Surgery.*

Heck, D. A., C. A. Melfi, L. A. Mamlin, B. P. Katz, D. S. Arthur, R. S. Dittus, and D. A. Freund. 1997. "Revision Rates Following Knee Replacement in the United States." Working Paper. Indiana University.

Hosenpud, J. D., T. J. Breen, E. B. Edwards, O. P. Daily, and L. G. Hunsicker. 1994. "The Effect of Transplant Center Volume on Cardiac Transplant Outcome: A Report of the United Network of Organ Sharing Scientific Registry." *Journal of the American Medical Association* 271 (23): 1844–49.

Hughes, R. G., D. W. Garnick, H. S. Luft, S. J. McPhee, and S. S. Hunt. 1988. "Hospital Volume and Patient Outcomes: The Case of Hip Fracture Patients." *Medical Care* 26 (11): 1057–67.

Iezzoni, L. I., J. Daley, T. Heeren, S. M. Foley, E. S. Fisher, C. Duncan, J. S. Hughes, and G. A. Coffman. 1994. "Identifying Complications of Care Using Administrative Data." *Medical Care* 32 (7): 700–15.

Katz, B. P., D. A. Freund, D. A. Heck, R. S. Dittus, J. E. Paul, J. Wright, P. Coyte, E. Holleman, and G. Hawker. 1996. "Demographic Variation in the Rate of Knee Replacement: A Multi-Year Analysis." *Health Services Research* 31 (1): 125–40.

Lubitz, R. M., R. S. Dittus, J. Paul, D. Z. Joycs, C. A. Melfi, R. L. Green, D. A. Freund, M. Fleming-Moran, D. A. Heck, and P. C. Coyte. 1995. "The Impact of Obesity on In-Hospital Outcomes Following Knee Replacement." Working Paper. Indiana University.

Luft, H. S. 1980. "The Relation Between Surgical Volume and Mortality: An Exploration of Causal Factors and Alternative Models." *Medical Care* 18 (9): 940–59.

Luft, H. S., J. P. Bunker, and A. C. Enthoven. 1979. "Should Operations Be Regionalized? The Empirical Relation between Surgical Volume and Mortality." *The New England Journal of Medicine* 301 (25): 1364–69.

Luft, H. S., S. S. Hunt, and S. C. Maerki. 1987. "The Volume-Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns?" *Health Services Research* 22 (2): 157–82.

Mitchell, J. B., T. Bubolz, J. E. Paul, C. L. Pashos, J. J. Escarce, L. H. Muhlbaier, J. M. Wiesman, W. W. Young, R. S. Epstein, and J. C. Javitt. 1994. "Using Medicare Claims for Outcomes Research." *Medical Care* 32 (7): JS38–JS51.

Norton, E. C., G. A. Zarkin, B. Calingaert, and C. J. Bradley. 1996. "The Effect of Maternal Substance Abuse on the Cost of Neonatal Care." *Inquiry* 33 (3): 247–57.

Norton, E. C., G. S. Bieler, S. T. Ennett, and G. A. Zarkin. 1996. "Analysis of Prevention Program Effectiveness with Clustered Data Using Generalized Estimating Equations." *Journal of Consulting and Clinical Psychology* 64 (5): 919–26.

Paul, J. E., C. A. Melfi, and L. A. Kalasinski. 1994. "Use of Large Databases for Outcomes Research." *Current Opinion in Orthopedics* 5 (1): 75–79.

Phillips, K. A., H. S. Luft, and J. L. Ritchie. 1995. "The Association of Hospital Volumes of Percutaneous Transluminal Coronary Angioplasty With Adverse Outcomes, Length of Stay, and Charges in California." *Medical Care* 33 (5): 502–14.

Romano, P. S., and H. S. Luft. 1992. "Getting the Most Out of Messy Data: Problems and Approaches for Dealing with Large Administrative Data Sets." In *Medical Effectiveness Research Data Methods,* edited by M. L. Grady and H. A. Schwartz. AHCPR Pub. No. 92–0056. Rockville, MD: Agency for Health Care Policy and Research, June.

Roos, N. P., C. D. Black, L. L. Roos, R. B. Tate, and K. C. Carriere. 1995. "A Population-Based Approach to Monitoring Adverse Outcomes of Medical Care." *Medical Care* 33 (2): 127–38.